Theoretical and Applied Genetics (2021) 134:1185–1200
https://doi.org/10.1007/s00122-020-03763-1

ORIGINAL ARTICLE

Quantitative trait mapping of plant architecture in two BC_1F_2 populations of *Sorghum Bicolor* × *S. halepense* and comparisons to two other sorghum populations

WenQian Kong[1] · Pheonah Nabukalu[2] · T. S. Cox[2] · Valorie H. Goff[1] · Jon S. Robertson[1] · Gary J. Pierce[1] · Cornelia Lemke[1] · Rosana Compton[1] · Andrew H. Paterson[1]

Received: 23 September 2020 / Accepted: 30 December 2020 / Published online: 9 January 2021

Abstract

Key message **Comparing populations derived, respectively, from polyploid *Sorghum halepense* and its progenitors improved knowledge of plant architecture and showed that *S. halepense* harbors genetic novelty of potential value for sorghum improvement**

Vegetative growth and the timing of the vegetative-to-reproductive transition are critical to a plant's fitness, directly and indirectly determining when and how a plant lives, grows and reproduces. We describe quantitative trait analysis of plant height and flowering time in the naturally occurring tetraploid ***Sorghum halepense***, using two novel BC_1F_2 populations totaling 246 genotypes derived from backcrossing two tetraploid ***Sorghum bicolor*** x ***S. halepense*** F_1 plants to a tetraploidized ***S. bicolor***. Phenotyping for two years each in Bogart, GA and Salina, KS allowed us to dissect variance into narrow-sense genetic (QTLs) and environmental components. In crosses with a common ***S. bicolor*** BTx623 parent, comparison of QTLs in ***S. halepense***, its rhizomatous progenitor ***S. propinquum*** and ***S. bicolor*** race guinea which is highly divergent from BTx623 permit inferences of loci at which new alleles have been associated with improvement of elite sorghums. The relative abundance of QTLs unique to the ***S. halepense*** populations may reflect its polyploidy and subsequent 'diploidization' processes often associated with the formation of genetic novelty, a possibility further supported by a high level of QTL polymorphism within sibling lines derived from a common ***S. halepense*** parent. An intriguing hypothesis for further investigation is that polyploidy of ***S. halepense*** following 96 million years of abstinence, coupled with natural selection during its spread to diverse environments across six continents, may provide a rich collection of novel alleles that offer potential opportunities for sorghum improvement.

Introduction

While the genetics of sorghum (*S. bicolor*) is relatively well studied, a wild relative with many intriguing properties has been largely overlooked. *Sorghum halepense* is a naturally formed polyploid that has spread from its suspected west Asian origin across six continents in post-Columbian times. Its establishment in the USA is probably typical of its spread to other continents, being introduced intentionally as a prospective forage and unintentionally as a contaminant of seedlots (McWhorter 1971), achieving the rare distinction of being both a noxious weed in 20 US states and an invasive species in 16 (Quinn et al. 2013). However, while sorghum largely remained confined to cultivation, *S. halepense* readily naturalized and has spread across much of North America, both to agricultural and non-agricultural habitats

---

Communicated by Jianbing Yan.

Supplementary Information The online version of this article (https://doi.org/10.1007/s00122-020-03763-1) contains supplementary material, which is available to authorized users.

✉ Andrew H. Paterson
paterson@uga.edu

[1] Plant Genome Mapping Laboratory, University of Georgia, 111 Riverbend Rd, Athens, GA 30602, USA

[2] The Land Institute, 2440 E Water Well Rd, Salina, KS 67401, USA

(Sezen et al. 2016)—suggesting capabilities for adaptation well beyond those of sorghum.

Modification of two important traits, plant height and flowering time, has been central to plant domestication and crop improvement. For example, the 'Green Revolution' was based largely upon short and lodging resistant plants suitable for increased fertilization and mechanical harvesting. Increasing demand for lignocellulosic biomass has motivated increasing plant height or developing dual purpose plants with both grain and biomass potential (Fernandez et al. 2009). Adjusting flowering time is important for hybrid development, for utilization of germplasm in non-native latitudes, and for adaptation to changing climatic conditions (Jung and Müller 2009).

One of the best-studied components of plant architecture is plant height, because of its importance, reliable phenotyping and relatively high heritability (Fernandez et al. 2009; Wang and Li 2006). However, it took nearly 40 years after the 'Green Revolution' for the underlying genes to be identified, semi-dwarf (*SD1*) in rice and reduced height (*Rht*) in wheat, both involved in gibberellin (GA) pathways (Peng et al. 1999; Sasaki et al. 2002; Wang et al. 2017). In addition to GA pathways, the brassinosteroid (BR) pathway is responsible for cell elongation by cell wall loosening (Kutschera and Wang 2012). Mutants of the BR pathway show dwarf phenotypes, while increased BR levels can increase plant size (Bishop 2003). Quantitative studies suggested that control of plant height is polygenic in maize (Ku et al. 2015; Peiffer et al. 2014).

In sorghum, a canonical model has suggested four general loci for plant height, *DW1- DW4* (Quinby and Karper 1954), but multiple quantitative studies (Brown et al. 2008; Hart et al. 2001; Lin et al. 1995; Morris et al. 2013; Murray et al. 2008; Ritter et al. 2008; Upadhyaya et al. 2012; Zhang et al. 2013; Zhang et al. 2015) support that at least six non-overlapping loci contribute to plant height (Zhang et al. 2015). To date, *Dw1* and *Dw3* have been cloned (Multani et al. 2003; Yamaguchi et al. 2016); *Dw2* is located in or near a large heterochromatin region on sorghum chromosome 6 (Brown et al. 2008; Cuevas et al. 2016; Higgins et al. 2014; Hilley et al. 2017; Murphy et al. 2011; Murray et al. 2009; Zhang et al. 2015); and *Dw4* may be near ~ 6.6 Mb on chromosome 6 (Morris et al. 2013), but strong signals on chromosome 4 have also been considered as indicators of *Dw4* (Li et al. 2015; Zhang et al. 2015). Additional loci with small effects responsible for plant height might explain residual variance.

Genetic manipulation of flowering time is important in crop domestication, and directly or indirectly influences plant architecture (Hill and Li 2016). In most Poaceae taxa, plants flowering late are taller, as flowering terminates apical growth. Indeed, many studies reported positive correlations between plant height and flowering time in sorghum and often discovered QTLs for these two traits located in the same general genomic regions (Brown et al. 2008; Higgins et al. 2014; Lin et al. 1995; Morris et al. 2013; Zhang et al. 2015). A total of six maturity genes controlling days to flowering have been denoted *Ma1–Ma6* (Brady 2006; Quinby 1966; Quinby and Karper 1945). An intriguing early candidate for *Ma1* on sorghum chromosome 6 (Murphy et al. 2011) has recently been found unlikely, and replaced by a more probable candidate supported by numerous lines of evidence (Cuevas et al. 2016). *Ma3* is proposed to be the phytochrome B on sorghum chromosome 1 (Childs et al. 1997; Yang et al. 2014a), while *Ma6* is hypothesized to be an ortholog of a rice grain number, plant height and heading date gene (*GHD7*) (Murphy et al. 2014), albeit needing further functional validation.

Sorghum provides many avenues to study traits related to plant architecture, thanks to rich genomic resources and a high-quality reference genome sequence (Paterson et al. 2009). In addition to genomic tools, the flexibility to make crosses between the five main sorghum races (bicolor, guinea, caudatum, durra and kafir), and with wild relatives such as *S. propinquum* and *S. halepense*, makes it particularly attractive to dissect and compare genetic components of plant architecture. In this paper, we describe a quantitative trait study of key components of plant architecture, specifically plant height and flowering time, in two half-sib tetraploid BC_1F_2 populations derived from crossing *Sorghum bicolor* BTx623 and *Sorghum halepense* Gypsum 9E. A two-year, two-environment phenotypic evaluation permits us to identify major effect and environment-specific QTLs. Quantitative trait loci (QTLs) discovered in these two novel populations are compared to those from two other diploid sorghum recombinant inbred line (RIL) populations sharing BTx623 as a common parent but sampling the breadth of the Sorghum genus, one a cross to *S. bicolor* IS3620C, and the other to *S. propinquum* (Kong et al. 2013). QTLs identified in this study and their comparison provide insight into evolution of morphological diversity in the Sorghum genus, are of practical use for marker-assisted breeding, and provide a foundation for molecular cloning and functional analysis.

Materials and methods

Genetic stocks

S. bicolor BTx623 × *S. halepense* G9E (SH-BC_1F_1 and SH-BC_1F_2): Two tetraploid F_1 hybrids, H4 and H6, derived from crossing *S. bicolor* BTx623 × *S. halepense* Gypsum 9E (G9E) were backcrossed to the tetraploidized recurrent parent, *S. bicolor* BTx623 to develop H4 and H6-derived BC_1F_1 mapping populations consisting of 141 and 105 genotyped individuals, respectively. BC_1F_2 rows derived from selfing single BC_1F_1 plants were planted on May 28th 2013 and May 9th 2014 at the University of Georgia Plant Science Farm, Watkinsville, GA, USA (33.87°, – 83.53°, Athens 2013 and Athens 2014 hereafter), and on Jun 3rd 2013, and Jun 17th 2014 at The Land Institute, Salina, KS, USA (38.77°, – 97.57°, Salina 2013 and Salina 2014 hereafter). Within each environment, there were two blocks and three subsamples for each genotype, in a completely randomized design.

RIL population of *S. bicolor* BTx623 × IS3620C (IS-RIL): This population comprised 393 $F_{7–8}$ RILs derived by selfing of a single F_2 plant from *S. bicolor* BTx623 × IS3620C, expanding a population previously described (Hart et al. 2001; Kong et al. 2000) and were described in Kong et al. (2018). This population was planted at the University of Georgia Plant Science Farm, Watkinsville, GA, USA on 10 May 2011 and 18 May 2012. Single 1.5-m plots of each RIL were machine planted in a completely randomized design. For each progeny line, we phenotyped two plants as subsamples.

RIL population of *S. bicolor* BTx623 and *S. propinquum* (PQ-RIL): This population comprised 161 RILs derived by selfing a single F_2 plant from *S. bicolor* BTx623 and *S. propinquum*, as described in Kong et al. (2013). The population was planted on May 20th 2009, May 28th 2010 and May 16th 2011 in a completely randomized design, transplanting (2009, 2011) or direct seeding (2010) five plants in a 1.5 m plot. For each progeny line, we phenotyped two plants as subsamples. Genetic components of vegetative branching patterns were elaborated in Kong et al. (2014).

Genotyping

Leaf samples of the SH-BC_1F_1 population were frozen at – 80°C and lyophilized for 48 h. Genomic DNA was extracted from the lyophilized leaf sample based on Aljanabi et al. (1999).

The GBS platform is a slightly modified version of Multiplex Shotgun Genotyping (MSG) (Andolfatto et al. 2011) combined with the Tassel GBS 5 analysis pipeline. Genotyping of the two SH-BC_1F_1 populations used Illumina HiSeq 2500, Rapid V2 kits that generate about 150 million reads of 100 base pair (bp) fragments per run with single-end sequencing. Details of SNP calling, marker filtering, and genetic map construction can be found in Kong et al. (2020). Genotyping of the IS-RIL population used an in-house Illumina MiSeq that generates up to 25 million reads of 150 base pair (bp) fragments per run with single-end sequencing. Description of the genetic map of the IS-RIL population can be found in Kong et al. (2018). The genetic map of the PQ-RIL population derived from *S. bicolor* BTx623 × *S. propinquum* was published in Kong et al. (2013).

Phenotyping

We evaluated plant height and flowering time in the SH-BC_1F_2 families with three subsamples in two fields in two years and at two locations. Flowering time (**FL**) was measured by recording when flower heads emerged for about 50% of the plants within a plot. Plant height (**PH**) was measured from the base to the tip of the main flower head. Phenotyping of plant height and flowering time in the IS-RIL population was consistent with our system applied to the PQ-RIL as described in Kong et al. (2014). The variance component method was used to calculate broad-sense heritability [$H = V_G/(V_G + V_{GE}/e + V_{residual}/er)$] where V_G is the variance estimate for genotype, V_E is the variance estimate for environment, e is the number of environments and r is the number of subsamples.

Single marker analysis

We screened informative markers to identify an appropriate subset for conducting single marker analysis based on the following procedure:

1. Select bi-allelic markers.
2. Select markers that are at least 100 bp apart since SNP markers within 100 bp are highly correlated and provide little additional information in genetic mapping populations in strong linkage disequilibrium.
3. Keep markers with missing data less than 60%.
4. Delete markers with minor allele frequency less than 0.02.
5. Calculate pair-wise recombination frequency, and remove markers that fail to show linkage to any other markers of < 20% recombination.

After filtering the raw dataset, we conducted one-way analysis of variance (ANOVA) for each phenotypic trait as the response variable with respect to each marker genotype for each population. A significance threshold of a *P* value of 0.001 was used. We also conducted ANOVA by pooling both SH-BC_1F_2 populations to increase statistical power while controlling population as a blocking factor. Statistical analyses used R (R Core Team 2016). The significant lists of SNPs for each trait in each sub-population were further condensed using hierarchical clustering with pair-wise recombination frequencies as the distance measures. We visualized the clustering of SNPs using heat maps with the R 'gplots' function (Warnes et al. 2016). Potential QTLs were determined if more than 4 SNPs were found within a cluster cut at height of 0.3 (30% recombination). Peak SNPs were chosen based on the smallest P-values.

Map-based QTL mapping

The interval mapping method (Lander and Botstein 1989) was conducted for each trait of interest in each population (H4 or H6) in each environment using R/qtl (Broman et al. 2003). Permutation tests (with $\alpha = 0.10$) suggest LOD scores of 2.9 and 3.1 for H4 and H6 populations, respectively. QTLs with LOD scores of 2.5 were listed. For QTLs showing multiple peaks or covering large genomic regions, we used the 'scantwo' function in R/qtl to test the hypothesis of two QTL models. After refining the QTL positions, an additive QTL model was employed to calculate the percentage of variance explained for each trait of each population in each environment. Then, a mixed model was fitted using all QTLs as fixed effects and the environment as a random effect to understand and partition the contribution of different QTL effects to the phenotype. We used a modified method to calculate R-squared (Nakagawa and Schielzeth 2013) for the mixed effect modeling.

Results

Summary statistics and heritability analysis

The average height of *S. halepense* G9E was 157 cm, much taller than the 98.7 cm of *S. bicolor*, albeit measured in a separate experiment in 2012 (2013–2014 data for G9E showed evidence of contamination). The tetraploid BTx623 parent was 36.7 cm taller than its diploid counterpart in 2013 ($t = 2.96$, $p = 0.0050$, Table S1a), and 28.1 cm taller in 2014 ($t = 4.91$, $p < 0.001$). The average height of the SH-BC$_1$F$_2$ progeny lines across all four environments is 250.2 cm, much taller than both parents, with plants grown in Salina averaging 42.7 cm taller than in Athens ($t = 17.27$, $p < 0.001$). The respective locations differed in opposite ways across years—average plant height in Athens was 31.3 cm shorter in 2013 than 2014 ($t = -9.38$, $p < 0.001$), but in Salina was 32.9 cm taller in 2013 than 2014 ($t = 10.88$, $p < 0.001$). The average height of the SH-BC$_1$F$_2$ population is the largest among the three sorghum populations, taller than the 151.2 cm of the IS-RILs ($t = 92.46$, $p < 0.001$) and the 100.6 cm of the PQ-RILs [($t = 40.49$, $p < 0.001$), (Table S1a, b, c)]. Broad-sense heritability estimates of plant height are relatively high and consistent among the three populations, at 72%, 78% and 77% for the SH-BC$_1$F$_2$, IS-RIL and PQ-RIL populations, respectively.

Days to flowering (**FL**) of *Sorghum halepense* G9E averaged 14 days earlier than that of BTx623 in 2012 (see explanation above). Tetraploid BTx623 plants flowered about 6.9 days later on average than diploid BTx623 ($t = 2.49$, $t = 0.019$), and this difference was larger in 2013 (8.2 days, $t = 3.25$, $p = 0.0061$) than 2014 (4.0 days, $t = 3.35$, $p = 0.0053$) (Table S1a). Progeny lines of SH-BC$_1$F$_2$ display large genetic variation, but their average **FL** was about 75.7 days, near the 76.4 day average of tetraploid BTx623. The average **FL** in Salina was about 4 days longer than in Athens ($t = 5.38$, $p < 0.001$). Within each environment, average **FL** in Athens was about 2.8 days later in 2013 than 2014 ($t = 2.70$, $p = 0.0071$); and in Salina was about 12.9 days later in 2013 than 2014 ($t = 14.42$, $p < 0.001$), a much larger difference than in Athens. Progenies of the BC$_1$F$_2$ population flowered an average of 16.9 days later than the IS-RIL ($t = 30.34$, $p < 0.001$), and 8.7 days later than the PQ-RIL (Table 1 b and 1c). Broad sense heritability estimates of days to flowering are relatively high, at 83.59%, 63.66% and 83.65% in the SH-BC$_1$F$_2$, PQ-RIL and IS-RIL populations, respectively.

Phenotypic values for **FL** are significantly correlated across environments, as are phenotypic values for **PH**.

Table 1 A mixed main effect model for plant height in the H4 derived population

	Sum Sq[1]	DF	Estimate[2]	*F* value	*P* value
qPH.3C.H4.1	22,923	1	13.21	27.95	1.806e−07***
qPH.6B.H4.1	26,783	1	16.66	32.66	1.815e−08***
qPH.6B.H4.2	37,628	1	22.78	45.89	3.278e−11***
qPH.7C.H4.1	18,812	1	15.02	22.94	2.161e−06***
qPH.7C.H4.2	10,347	1	10.90	12.62	0.0004154***

Source	Variance	SD
Random effects		
ENV	836.1	28.92
Residual	820.0	28.64

Sum sq sum of squares, *DF* degrees of freedom, *ENV* environment, *SD* standard deviation

[1] Sum of squares of a single QTL in the mixed effect model

[2] Estimated effects of allele substitution

Further, **FL** and **PH** are also positively and significantly correlated with one another (Fig. 1), indicating that late flowering individuals are generally taller than early flowering ones.

QTL analysis

In addition to the conventional interval mapping method (Lander and Botstein 1989), we performed single marker analysis for each trait with respect to each of the 5148 filtered SNP markers, with the order of the markers based on the published sorghum genome sequence version 1.4 (Paterson et al. 2009). An example of a heat map with hierarchical clustering was provided to visualize the relationship among significant markers and to determine the number of potential QTLs (Fig. S1). This is a greedy method but nevertheless permits us to analyze pooled data from the two BC_1F_2 populations and to visualize the relationship between blocks of potential QTLs without constructing genetic maps.

Plant height

We detected a total of 11 QTLs for **PH** in the H4-derived population by interval mapping, with six, qPH3C.H4.1, qPH6B.H4.1, qPH7C.H4.1, qPH7C.H4.2, qPH7D.H4.1 and qPH9B.H4.1 significant in more than one environment (Fig. 2, Table S2). Allele effects of the detected QTLs are all positive, indicating that *S. halepense* alleles increase **PH**. A total of five QTLs, qPH3C.H4.1, qPH6B.H4.1, qPH6B.H4.2, qPH7C.H4.2 and qPH7C.H4.3, were selected using backward selection ($p < 0.05$) to fit a mixed main effect model for **PH** together with the environmental factor as the random effect (Table 1). The model with five QTLs and one environmental factor explains 63.2% of the total variance, with

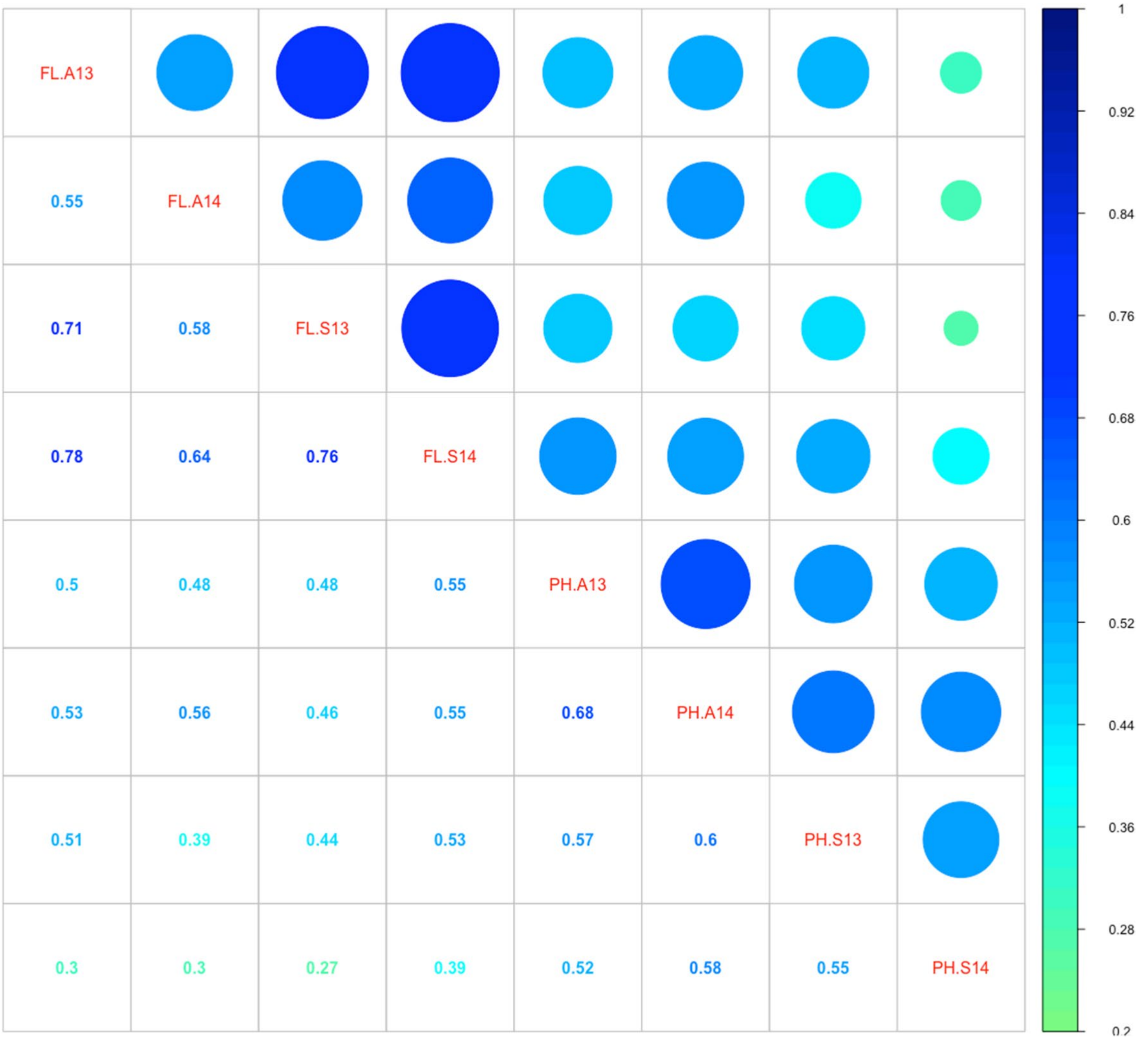

Fig. 1 Correlation coefficients between days to flowering (FL) and plant height (PH) in the BC_1F_2 populations in four environments. All correlation coefficients are significant at an alpha level of 0.001

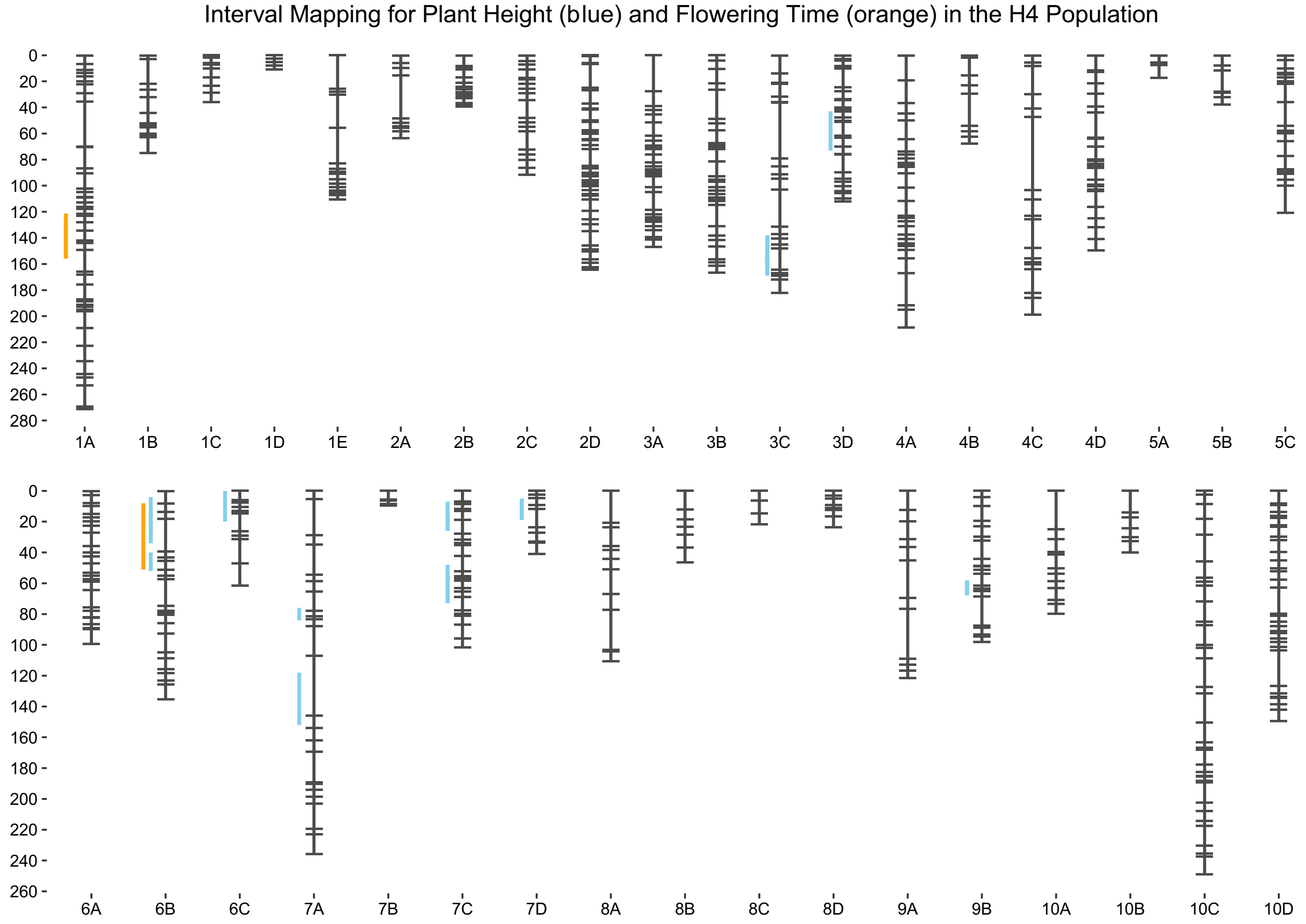

Fig. 2 Interval mapping result of the H4 population

the QTL factor (fixed) explaining 24.3% of the total variance based on a modified method for calculating R-squared in the mixed model (Nakagawa and Schielzeth 2013). The random effects of the four environments from the mixed model are − 34.20, − 3.29, 35.47, 2.01 cm for Athens 2013, Athens 2014, Salina 2013 and Salina 2014, respectively, suggesting that the progeny plants were tallest in Salina 2013 and shortest in Athens 2013 among the four environments.

In the H6-derived population, we detected a total of 20 **PH** QTLs by interval mapping, with 11 significant in more than one environment (Fig. 3 and Table S2). Sixteen QTLs show positive allele substitution effects, with five, qPH4A.H6.1, qPH.4C.H6.1, qPH.4C.H6.2, qPH.8A.H6.1 and qPH.9C.H6.1 showing negative allele substitution effects, indicating that *S. halepense* alleles decrease plant height at these loci. qPH4C.H6.1, qPH8A.H6.1 and qPH9C.H6.1 might be reciprocal alleles (i.e., different homologs) to qPH4D.H6.1, qPH8B.H6.1 and qPH9A.H6.1 or qPH9B.H6.1 based on comparing their physical positions. A mixed effect model using backward elimination ($p = 0.05$) with the environment as a random effect selected a total of nine QTLs, qPH1A.H6.1, qPH.1C.H6.1, qPH.2B.H6.1, qPH.3B.H6.1, qPH.4A.H6.1, qPH.4D.H6.1, qPH.6B.H6.1, qPH.7D.H6.1 and qPH.9A.H6.1 (Table 2), with only one QTL (qPH.4D.H6.1) showing negative allele effect of *S. halepense*. This provides some evidence for our hypothesis of reciprocal QTLs, since only one QTL in each pair is significant in the mixed effect model. QTL factors (fixed) of this model explain about 22.4% of the total variance, while inclusion of the environmental factor (random) explains about 71.9% of the total variance. The random effects of the four environments from the mixed model are − 28.73, − 5.86, 28.77 and 5.82 for Athens 2013, Athens 2014, Salina 2013 and Salina 2014, respectively, suggesting again that the progeny plants were tallest in Salina 2013 and shortest in Athens 2013.

We detected a total of 243, 239, 89 and 78 significant SNP markers ($p < 10^{-3}$) for **PH** in Athens 2013, Athens 2014, Salina 2013 and Salina 2014, respectively, with 34 markers significant in all four environments (Fig. 4 and Fig. S2). In the H4 derived population, we inferred a total of eight QTLs on chromosomes 2, 3, 5, 6 (2), 7 (2), 9, all

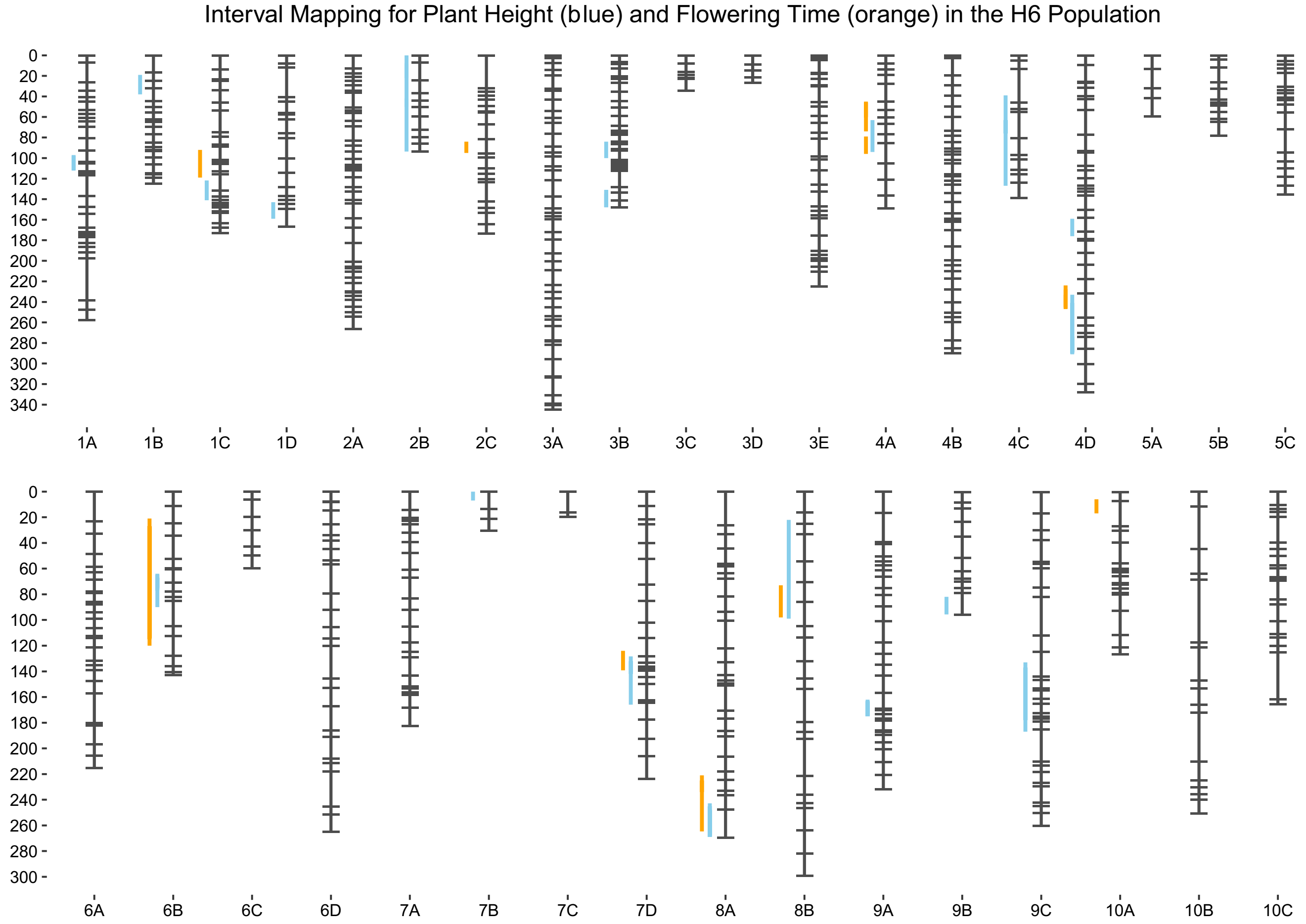

Fig. 3 Interval mapping result of the H6 population

increasing **PH,** with one QTL, qPH2.1.H4 newly detected. In the H6-derived population, we detected a total of 14 QTLs including five pairs of QTLs (reciprocal QTLs) on the same chromosome but with different effects, possibly coming from homologs in this population; single QTLs on each of the remaining chromosomes except chr. 5. A total of two QTLs, qPH10.1.H6 and qPH10.2.H6 were newly detected from the single-marker analysis in the H6-derived population. Three previously mapped genes, *Dw1*, *Dw2* and *Dw3* (Morris et al. 2013; Multani et al. 2003; Yamaguchi et al. 2016), were tentatively re-identified in our single marker analysis, corresponding to the 'skyline' signals on the long arms of chromosomes 9, 6 and 7, indicating that *S. halepense* has wild-type alleles that increase plant height at those loci. Many small effect signals were found across the genome (Table S2 and S3), which support findings that the genetic control of plant height in sorghum is quantitative with many more loci involved (Brown et al. 2008; Hart et al. 2001; Lin et al. 1995; Murray et al. 2008; Ritter et al. 2008; Upadhyaya et al. 2012; Zhang et al. 2015) than the canonical model based on four loci, *DW1-DW4* (Quinby and Karper 1954).

A total of four general **PH** QTL regions for the two SH-BC_1F_2 populations overlap, based on aligning their physical positions on the sorghum genome, including those regions near the *DW1* (qPH9B.H4.1 and qPH9B.H6.1) (Hilley et al. 2016; Yamaguchi et al. 2016), *DW2* (qPH6B.H4.2 and qPH.6B.H6.1) and *DW3* loci (qPH.7C.H4.2, qPH7D.H4.1, qPH7B.H6.1 and qPH7D.H6.1) (Multani et al. 2003) and a fourth region on chromosome 3 (qPH3C.H4.1 and qPH3B.H6.1) at approximately 68–69 Mb. More QTLs are detected in the H6 than the H4 population, possibly because the short-day alleles on chromosome 6 have much less effect in the H6 than the H4 population.

Days to flowering

A total of three flowering time QTLs, qFL.1A.H4.1, qFL6B.H4.1 and qFL6B.H4.2, were detected by the interval mapping method from the H4-derived population, all significant in more than one environment (Fig. 2 and Table S4).

Table 2 A mixed main effect model for plant height in the H6 population

	Sum Sq	DF	Estimates	*F* value	*P* value
qPH.1A.H6.1	8239	1	– 19.53	14.28	0.0001814***
qPH.1C.H6.1	16,008	1	16.20	27.75	2.267e–07***
qPH.2B.H6.2	4335	1	7.56	7.51	0.0063978**
qPH.3B.H6.1	6518	1	9.87	11.30	0.0008510***
qPH.4A.H6.1	7605	1	– 13.82	13.18	0.0003195***
qPH.4C.H6.2	3710	1	– 7.91	6.43	0.0115983*
qPH.4D.H6.1	10,734	1	17.55	18.61	2.030e–05***
qPH.6B.H6.1	47,525	1	25.83	82.38	< 2.2e–16***
qPH.7D.H6.1	5812	1	– 12.35	10.07	0.0016203**
qPH.8B.H6.1	1841	1	– 5.32	3.19	0.0748162
qPH.9A.H6.1	5177	1	14.66	8.97	0.0029092**

Source	Variance	SD
Random effects		
ENV	1065.8	32.65
Residual	576.9	24.02

Sum sq sum of squares, *DF* degrees of freedom, *ENV* environment, *SD* standard deviation

[1] Sum of squares of a single QTL in the mixed effect model

[2] Estimated effects of allele substitution

Sorghum halepense alleles of both QTLs from chromosome 6B (qFL6B.H4.1 and qFL6B.H4.2) delay flowering while *S. halepense* alleles of qFL.1A.H4.1 accelerate flowering. A log likelihood plot (Fig. S3) suggests two possible QTLs on chromosome 6B, and a 'scantwo' analysis in R/qtl for all four environments favors two QTLs on chromosome 6B, with likelihood peaks at 6–8 cM and 46–49 cM, corresponding to 0.9 and 40–43 Mb in physical location. An average LOD score is 3.58 when comparing the full model to the single QTL model, and 3.02 when comparing the additive model to the single QTL model (data not shown). We obtained a mixed effect model adding all three QTLs, qFL.1A.H4.1, qFL6B.H4.1 and qFL6B.H4.2, as the fixed effects significant at an alpha level of 0.05 and the environment as a random effect (Table 3). QTL by environment interactions are not significant for **FL** in H4 for this analysis. The fixed effect (QTL) explains about 28.94% of the total variance, while the model including both fixed and random effects explains about 49.74% of the variance. This indicates that flowering time varies substantially between years, and the random effects of environments are – 1.86, – 3.97, 13.29, – 7.45 days, suggesting that **FL** in Salina 2013 took much longer than any other environment (also see Table 1a).

We detected a total of 10 QTLs for **FL** by interval mapping in the H6 population with 4, qFL4A.H6.1, qFL4D.H6.1, qFL6B.H6.1, qFL8A.H6.1 significant in more than one environment (Fig. 3 and Table S4). *Sorghum halepense* alleles accelerate flowering at qFL.4A.H6.1, qFL.4A.H6.2, qFL.8A.H6.1 and qFL.10A.H6.1, and delay flowering at the rest of the QTLs. We suspect that there are more than one QTL on chromosome 6B (as was found in the H6 population); however the 'scantwo' result does not formally support a two-QTL model (data not shown). We obtained a mixed effect model with a total of 8 QTLs, qFL.1C.H6.1, qFL.2C.H6.1, qFL.4A.H6.1, qFL.4A.H6.2, qFL.6B.H6.1, qFL.6B.H6.2, qFL.8A.H6.1, qFL.10A.H6.1 (Table 4), as the fixed effects significant at an alpha level of 0.05 and environment as the random effect. The QTL (fixed) factor explains about 25.77% of the total phenotypic variance, while the QTLs (fixed) and the environment (random) factors collectively explain about 52.34% of the total phenotypic variance. The random effect estimates from the mixed model are 0.49, – 5.85, 10.71, – 5.35 for the four environments, suggesting again that **FL** is much later in Salina 2013 than any other environment.

We detected totals of 118, 84, 94 and 94 significant SNP markers ($p < 10^{-3}$) for **FL** in Athens 2013, Athens 2014, Salina 2013 and Salina 2014, respectively, with 54 SNP markers significant in all four environments (Fig. 5 and Fig. S4). The overall distribution of pooled data for **FL** suggested two major peaks on sorghum chromosome 6, with peak markers being S6_941772 and S6_42153422; and one peak on sorghum chromosome 1 at SNP marker S1_20362820 (Fig. 3). Average *S. halepense* allele effects delay flowering on chromosome 6 but accelerate flowering on chromosome 1. The hierarchical clustering of significant markers suggested a total of four possible QTLs, one on chromosome 1 and three on chromosome 6 in the H4-derived population

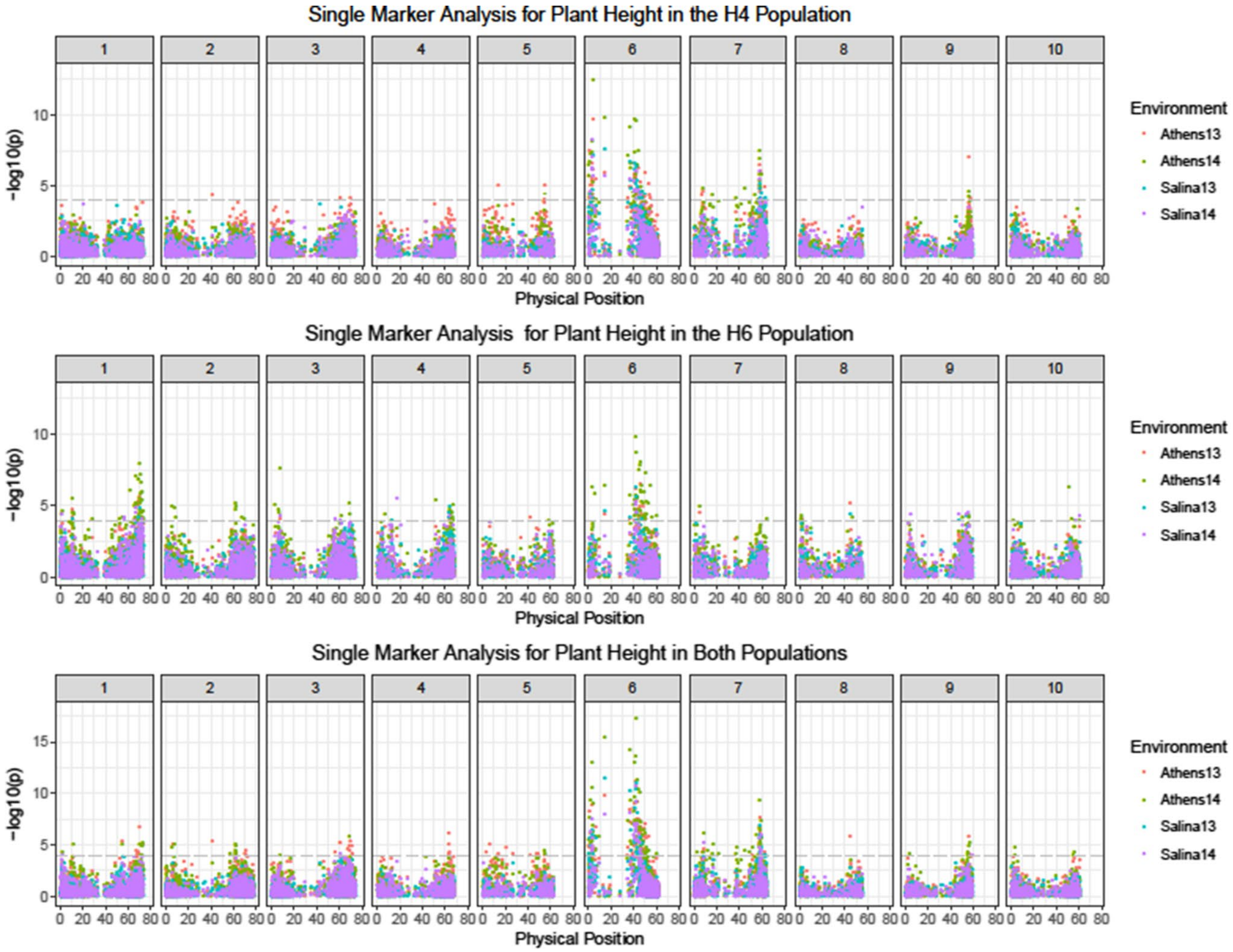

Fig. 4 Single marker analysis of plant height in the H4, H6 -derived and pooled BC_1F_2 populations. Physical positions are in Mb

Table 3 A mixed effect model for days to flowering (**FL**) in the H4 derived population

	Sum Sq	DF	Estimates	*F* value	*P* value
qFL.1A.H4.1	1162.8	1	− 4.1322	13.642	0.0002435***
qFL.6B.H4.1	5320.1	1	7.3288	62.416	1.554e−14***
qFL.6B.H4.2	6281.1	1	9.1940	73.69	< 2.2e−16***
Source			Variance		SD
Random effects					
Environment			34.94		5.911
Residual			84.24		9.232

Sum sq sum of squares, *DF* degrees of freedom, *ENV* environment, *SD* standard deviation

[1] Sum of squares of a single QTL in the mixed effect model

[2] Estimated effects of allele substitution

(Table S5), all overlapping with QTLs detected from the interval mapping. Allele substitution effects are negative on chromosome 1 but positive on chromosome 6. The three QTLs on chromosome 6 overlap based on their physical positions, therefore might be reciprocal QTLs. A total of seven possible QTLs were significant in the H6-derived

Table 4 A mixed effect model for days to flowering in the H6 derived population

	Sum Sq	DF	Estimates	*F* value	*P* value
qFL.1C.H6.1	502.20	1	2.5103	8.334	0.004100**
qFL.2C.H6.1	643.56	1	5.6277	10.680	0.001176**
qFL.4A.H6.1	1038.64	1	− 3.9864	17.236	4.030e−05***
qFL.4A.H6.2	405.95	1	−3.2681	6.737	0.009789**
qFL.6B.H6.1	1230.46	1	4.9764	20.420	8.184e−06***
qFL.6B.H6.2	355.87	1	2.5166	5.906	0.015527*
qFL.8A.H6.1	589.51	1	−3.2735	9.783	0.001889**
qFL.10A.H6.1	178.79	1	− 2.1279	2.967	0.085747

Source	Variance	SD
Random effects		
ENV	33.27	5.768
Residual	60.26	7.763

Sum sq sum of squares, *DF* degrees of freedom, *ENV* environment, *SD* standard deviation

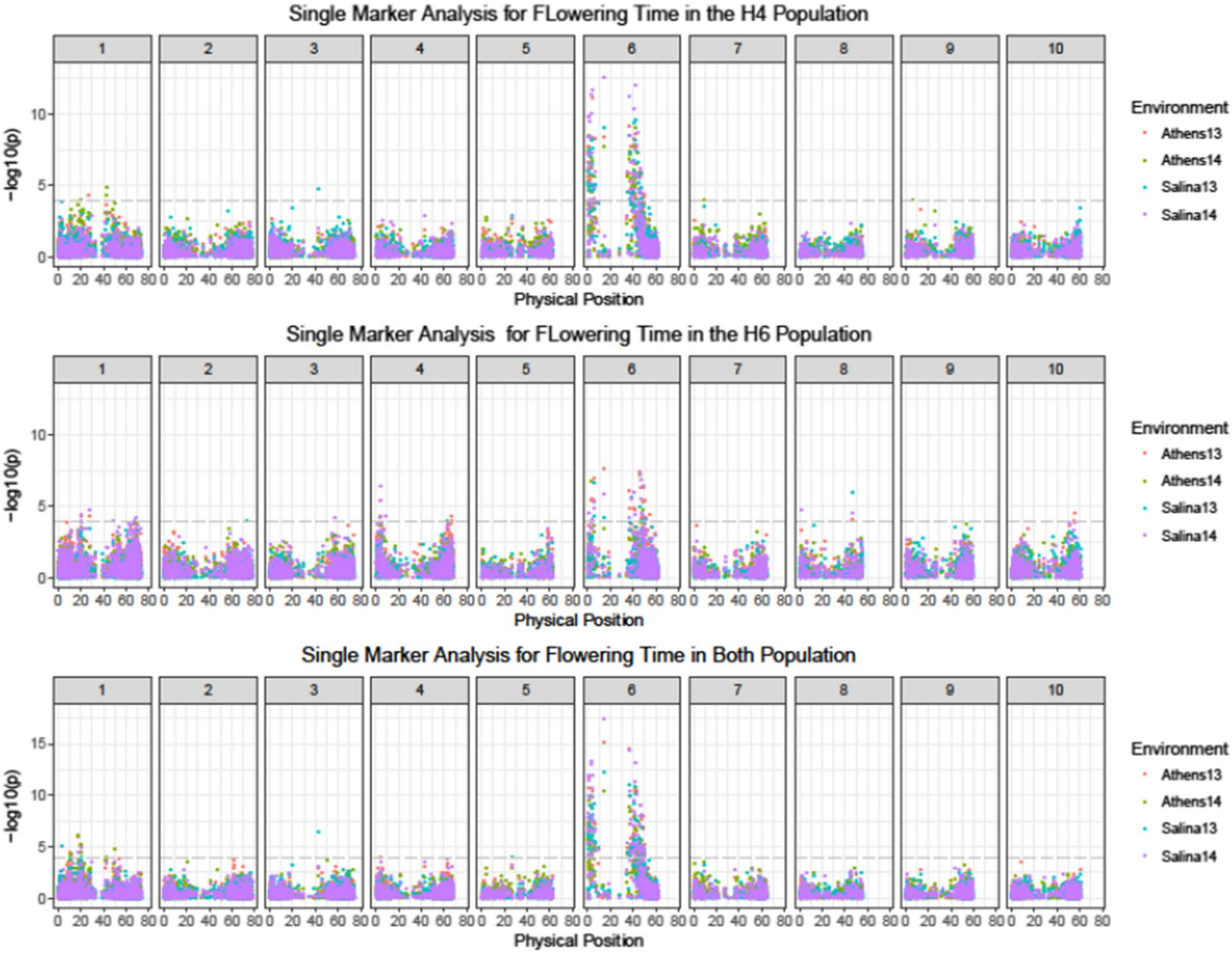

Fig. 5 Single marker analysis for days to flowering in the H4-, H6-derived and pooled BC_1F_2 populations. Physical positions are in Mb

population, distributed on chromosomes 1 (2), 4 (2), 6 (2), and 10, with two QTLs, qFL1.1.H6 and qFL10.1.H6 newly detected for the single marker analysis. It is interesting that chromosomes 1 and 4 contain QTLs both accelerating and delaying flowering time, and these are probably not reciprocal QTLs based on their distant physical positions.

In partial summary, QTL results for **FL** reveal large differences between the two SH-BC$_1$F$_2$ populations. The H4 population seems to be dominated by two QTLs on linkage group 6B with only one other QTL detected on linkage group 1A. The QTLs on linkage group 6B in the H6-derived population show more subtle effects and explain less phenotypic variance, perhaps contributing to our ability to detect nine more QTLs on other linkage groups. Both populations showed QTL effects in both directions, which might explain the transgressive segregation for flowering time observed in the SH-BC$_1$F$_2$ population. It also suggests that while *S. halepense* itself flowers rapidly, it can nonetheless contain late flowering alleles that are unmasked in segregating populations. However, the *S. halepense* QTL allele contributing to early flowering in the H4 population, qFL.1A.H4.1, is different from those in the H6 population, which are on chromosomes 4 and 8. We detected multiple QTLs on sorghum chromosome 6 which presumably harbors two flowering genes, with qFL.6B.H4.2 and qFL6B.H6.1 mapping near the location of *Ma1* (Cuevas et al. 2016; Murphy et al. 2011) and qFL6B.H4.1 mapping near *Ma6* (Murphy et al. 2014).

QTL correspondence across traits in the BC$_1$F$_2$ population

QTL correspondence among traits may be related to genes with pleiotropic effects, or multiple closely-linked QTLs responsible for different traits. **PH** and **FL** in the two BC$_1$F$_2$ populations are significantly correlated (Fig. 1), indicating that some genomic regions might affect both traits. In the H4-derived population, two QTLs conferring **FL** on chromosome 6B (qFL.6B.H4.1 and qFL6B.H4.2) overlap with two plant height QTLs (qPH.6B.H4.1 and qPH.6B.H4.2). The correspondence of qFL6B.H4.2 and qPH6B.H4.2 may reflect the well-known close linkage between the *Ma1* and *Dw2* genes (Cuevas et al. 2016; Lin et al. 1995; Morris et al. 2013), and qFL6B.H4.1 and qPH.6B.H4. might suggest a similar case, possibly *Ma6* and *Dw4*. In the H6-derived population, we detected a total of seven regions showing QTL correspondence between days to flowering and plant height on linkage groups 1C, 4A, 4D, 6B, 7D and 8A based on the interval based mapping, and one additional pair on chromosome 10 based on single marker analysis (Tables S2 and S4). Interestingly, one pair of overlapping QTLs, qFL.4A.H6.1 and qPH.4A.H6.1, show opposite effects compared to qFL.4D.H6.1 and qPH.4D.H6.1, with the first pair of **FL** and **PH** QTLs delaying flowering time and decreasing plant height and the second pair of **FL** and **PH** QTLs accelerating flowering time and increasing plant height.

Comparison to two other sorghum populations

We compared the QTL results from the SH-BC$_1$F$_2$ populations with two other sorghum populations sharing BTx623 as a common parent, specifically one RIL population (IS-RIL) derived from crossing *S. bicolor* BTx623 × IS3620C, and another RIL population derived from crossing *S. bicolor* BTx623 × *S. propinquum* (PQ-RIL) (Table S6 and S7, and Fig. 6). For **PH**, three general genomic regions conferring *Dw1*, *Dw2* and *Dw3* on chromosomes 7, 6, and 9, respectively, overlap between at least two populations with all three showing significant signals at the *Dw1* region on chromosome 7 and the *Dw2* and the *Dw3* regions significant in both SH-BC$_1$F$_2$ and IS-RIL populations (Table 5). Two additional PH QTL regions overlap on chromosomes 2 and 10 between the SBSH-BC$_1$F$_2$ and IS-RIL populations (Table 5).

A total of five genomic regions, on sorghum chromosomes 1, 4 (2), 6 and 10, showed significant association with **FL** in more than one population (Table 6 and Fig. 6). Both IS-RIL and PQ-RIL populations lack the short-day *Ma1* allele on chromosome 6 (IS3620c is a 'converted' sorghum, and the PQ-RIL population was advanced under day-neutral photoperiod in Lubbock TX), and accordingly we find no QTL correspondence in this region. Overlapping regions on chromosome 1, qFL.1A.H4.1 and qFL1.1 from ISRIL, might correspond to the sorghum *Ehd1* gene, which is thought to activate *Ft* expression (Doi et al. 2004; Murphy et al. 2011). Interestingly, the BTx623 allele for qFL.1A.H4.1 delays flowering in the BC$_1$F$_2$ population while that for qFL1.1 accelerates flowering in the IS-RIL population, again suggesting three alleles with *S. halepense* conferring the most rapid flowering. In addition, qFL4A.H6.1 and qFL4D.H6.1 on sorghum chromosome 4 express different allele effects, corresponding to qFL4.1 from PQ-RIL and qFL4.1 from IS-RIL, respectively. BTx623 alleles of qFL4A.H6.1 delay flowering time in the BC$_1$F$_2$ population but of qFL4.1 from PQ-RIL accelerate PQRIL flowering, again consistent with three *S. halepense* alleles conferring the most rapid flowering. However, for qFL4D.H6.1 and qFL4.1; and qFL10.H6.1 and qFL10.1 from SBSH-BC$_1$F$_2$ and IS-RIL, respectively, BTx623 alleles confer early flowering (Table 6) in both cases.

A total of 7 and 13 **PH** QTLs were unique to the H4 and H6 derived SBSH BC$_1$F$_2$ populations, respectively (Table 5). In addition, a total of six flowering QTLs, qFL1C.H6.1, qFL2C.H6.1, qFL7D.H6.1, qFL8A.H6.1, qFL8B.H6.1 and qFL10A.H6.1, were unique to the H6-derived BC$_1$F$_2$ population, indicating that those QTL alleles may have arisen during the radiation of *S. halepense* (Table 6). The fact that we detect more **PH** than

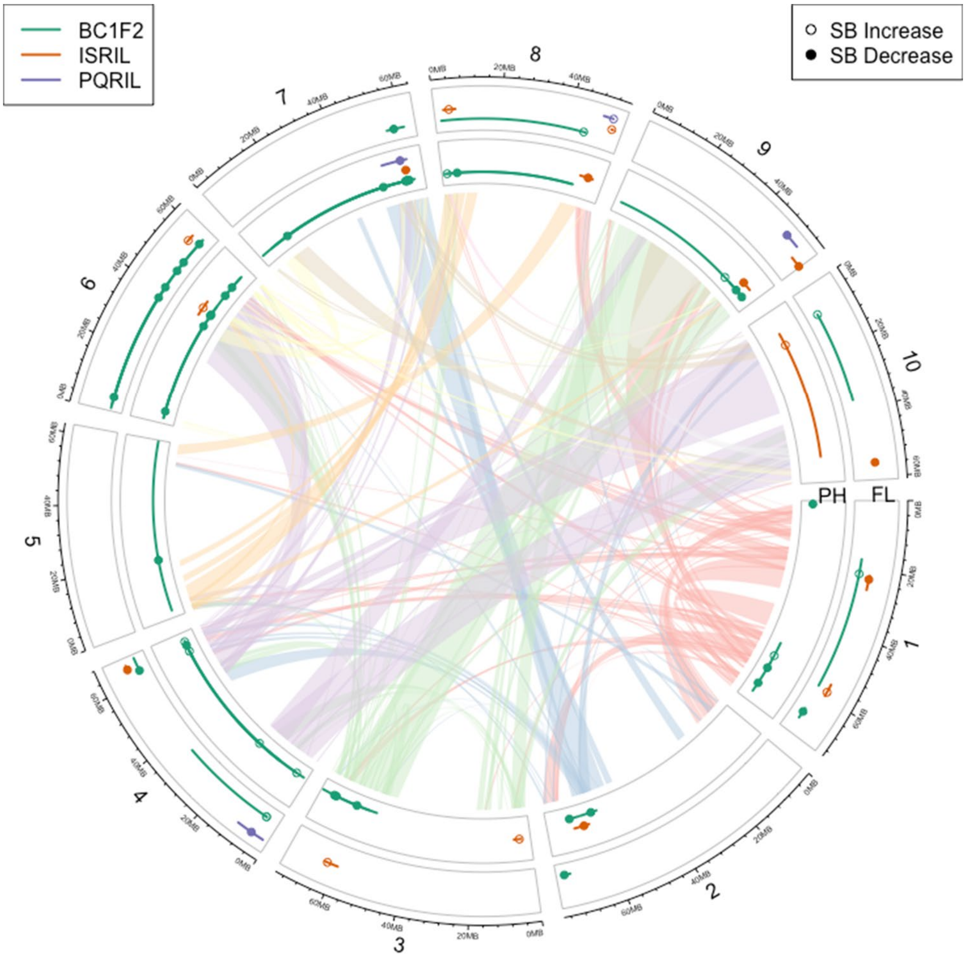

Fig. 6 QTL correspondence plot. QTL locations were converted into their physical positions. Links are genome duplication event in sorghum (Lee et al. 2013)

FL QTL, and that more **PH** than **FL** QTL are unique to the SBSH cross, may suggest that the genetic control of sorghum **FL** may be more conserved than that of **PH**. A total of three flowering QTLs, qFL1.2, qFL3.1, qFL8.1, and two plant height QTLs, qPH3.1 and qPH8.1 are unique to the IS3620C population, (Table S5), while none of the PQ-RIL QTLs are unique.

Discussion

The present study adds several new dimensions to knowledge of the *Sorghum* genus. First, it provides early insight into trait control and QTL polymorphism in *S. halepense*, one of the world's most important agricultural weeds and also an invasive plant now distributed over six continents. Second, the comparison of *S. halepense*, its progenitor *S. propinquum*, and a divergent form of *S. bicolor*, each crossed to the *S. bicolor* genotype from which the reference genome is derived, provides insight into the extent of genetic novelty that may have been associated with the evolution of polyploid *S. halepense* following an estimated 96 million years of abstinence from polyploidy in the sorghum lineage (Wang et al. 2015).

QTL allele polymorphism in S. halepense

The properties of the two SH-BC$_1$F$_2$ populations are quite different, reflecting a remarkably high level of QTL allele polymorphism in *S. halepense*—noting that the two F$_1$ source plants were obtained by crossing *S. bicolor* inbred line BTx623 to *S. halepense* accession Gypsum 9E, indicating a high level of polymorphism in the latter. Much of this difference appears to relate to the very strong effect of the chromosome 6 flowering genes in the H4-derived BC$_1$F$_2$ population, putatively *Ma1* and *Ma6* (see below). The striking effect of the chromosome 6 flowering genes in the H4-derived BC$_1$F$_2$ population might mask small effect flowering QTLs, accounting for the much larger number we found in the H6-derived BC$_1$F$_2$ population.

Table 5 Plant height (PH) QTL comparisons in SBSH BC1F2, IS-RIL and PQ-RIL population

QTL	IS-RIL	PQ-RIL	Inclusion of Genes
qPH2.H4.1 (+)			
qPH.3C.H4.1 (+)			
qPH.3D.H4.1 (+)			
qPH.5C.H4.1 (+)			
qPH.6B.H4.1 (+)			
qPH.6B.H4.2 (+)	qPH6.1 (−)		*Dw2*
qPH.6C.H4.1 (+)			
qPH.7A.H4.1 (+)			
qPH.7A.H4.2 (+)	qPH7.1 (+)	qPH7.1 (+)	*Dw3*
qPH.7C.H4.1 (+)			
qPH.7C.H4.2 (+)	qPH7.1 (+)	qPH7.1 (+)	*Dw3*
qPH.7D.H4.1 (+)	qPH7.1 (+)	qPH7.1 (+)	*Dw3*
qPH.9B.H4.1 (+)	qPH9.1 (+)		*Dw1*
qPH.1A.H6.1 (+)			
qPH.1B.H6.1 (+)			
qPH.1C.H6.1 (+)			
qPH.1D.H6.1 (−)			
qPH.2B.H6.1 (+)	qPH2.1 (+)		
qPH.2B.H6.2 (+)	qPH2.1 (+)		
qPH.3B.H6.1 (+)			
qPH.3B.H6.2 (+)			
qPH.4A.H6.1 (−)			
qPH.4C.H6.1 (−)			
qPH.4C.H6.2 (−)			
qPH.4D.H6.1 (+)			
qPH.6B.H6.1 (+)	qPH6.1 (−)		*Dw2*
qPH.7B.H6.1 (+)	qPH7.1 (+)	qPH7.1 (+)	*Dw3*
qPH.7D.H6.1 (+)	qPH7.1 (+)	qPH7.1 (+)	*Dw3*
qPH.8A.H6.1 (−)			
qPH.8B.H6.1 (+)			
qPH.9A.H6.1 (+)	qPH9.1 (+)		*Dw1*
qPH.9B.H6.1 (+)	qPH9.1 (+)		*Dw1*
qPH.9C.H6.1 (−)	qPH9.1 (+)		*Dw1*
qPH10.1.H6 (+)	qPH10.1 (−)		
qPH10.2.H6 (−)	qPH10.1 (−)		

Table 6 Flowering time (FL) comparisons in SBSH-BC1F2, IS-RIL and PQ-RIL population

QTLname	ISRIL	PQRIL	Inclusion of genes
qFL.1A.H4.1 (−)	qFL1.1 (+)		*Ehd1*
qFL.6B.H4.1 (+)			*Ma6*
qFL.6B.H4.1 (+)			*Ma6*
qFL.6B.H4.2 (+)			*Ma1*
qFL1.H6.1 (−)	qFL1.1 (+)		*Ehd1*
qFL.1C.H6.1 (+)			
qFL.2C.H6.1 (+)			
qFL.4A.H6.1 (−)		qFL4.1 (+)	
qFL.4A.H6.2 (−)		qFL4.1 (+)	
qFL.4D.H6.1 (+)	qFL4.1 (+)		
qFL.6B.H6.1 (+)			*Ma1*
qFL.6B.H6.2 (+)			*SbTFL1-1*
qFL.7D.H6.1 (+)			
qFL.8A.H6.1 (−)			
qFL.8B.H6.1 (+)			
qFL.10A.H6.1 (−)			*CO/SbMFT2*
qFL10.H6.1 (+)	qFL10.1 (+)		

SMA vs QTL mapping

We proposed a novel and fast method to visualize and characterize relationships of significant SNPs with single marker analysis and hierarchical clustering. Constructing genetic maps in these two largely auto-tetraploid populations derived from a heterozygous parent is relatively labor-intensive and challenging for accurate genotyping calling, requiring a high depth of coverage to call heterozygosity and separate linkage groups. The single-marker analysis method is an attractive alternative to identify significant SNPs associated with traits of interest. However, SNP signals tend to be dispersed along the 10 sorghum chromosomes in this study due to doubled-ploidy of *S. halepense*. Re-grouping the significant SNP signals is especially useful in categorizing SNPs with different signs of effect, separating QTLs from different homologs, and visualizing some small chromosomal differences between *S. bicolor* and *S. halepense*. Single marker analysis may also detect additional potential QTL signals that eluded interval mapping, especially where portions of chromosomes are not included in the genetic map.

We detected a total of three QTLs for PH and two for FL by single marker analysis in addition to those found by interval-based mapping, and the majority of QTLs detected by the single marker analysis found their counterparts in interval mapping. This result suggests that our analytical method functioned reasonably well in detecting and resolving the relationships of QTLs.

QTL mapping

In this experiment, we conducted QTL analysis of two traits with relatively high heritability estimates, **PH** and **FL,** finding many QTLs significant in multiple environments (Table S2–S5). Despite high heritability estimates, environmental factors still explain large portions of phenotypic variance in mixed effect models (Tables 1, 2, 3, 4).

The largest effect QTL for flowering time is on chromosome 6, and is especially pronounced in the H4 population where each of two chr. 6 QTLs have LOD scores greater than 10. Those two QTLs, qFL6B.H4.1 and qFL6B.H4.2

presumably correspond to the sorghum *Ma6* and *Ma1* genes (Brady 2006; Cuevas et al. 2016; Murphy et al. 2011). *Ma1* in particular is of great interest—tacitly assuming that *S. halepense* formed from progenitors resembling the wild sorghums of today, one would postulate that it had four (!) copies of the dominant *Ma1* allele conferring short-day flowering. In principle, this would delay flowering until day lengths drop below 12.5 h, during September in the latitude of the study site. However, Gypsum 9E, and indeed an entire diversity panel of 599 *S. halepense* accessions from across the USA, flowered throughout the long days of the temperate summer. Moreover, *S. halepense* flowered ~ 14 days earlier than day-neutral *S. bicolor* BTx623. The difference in flowering time between the two SH-BC$_1$F$_2$ populations appears to suggest either polymorphism for *Ma1* itself, or for some interacting factor that permits day-neutral flowering, a question for further study.

Other than the two flowering QTLs on chromosome 6, only one QTL, qFL1A.H4.1, was significant in the H4-derived population, suggesting that the large effect QTLs mask the effects of QTLs on other chromosomes. QTLs on linkage group 6B have smaller LOD values in the H6 than the H4-derived population, in which an additional 12 flowering QTLs were found. Across the two populations, a total of five QTLs, qFL.1A.H4.1 (overlapping with qFL1.H6.1 from single marker analysis), qFL.4A.H6.1, qFL.4A.H6.2, qFL8A.H6.1 and qFL.10A.H6.1 have *S. halepense* alleles accelerating flowering time. Thus, *S. halepense* harbors both alleles delaying flowering and alleles that accelerate it, accounting for the transgressive segregation of **FL** in progeny populations and potentially offering alleles that may be useful in breeding sorghum for specific environments.

In addition to re-identification of previously detected height genes *Dw1-Dw3*, we detected 7 and 17 more QTLs for plant height in the H4 and H6-derived populations, respectively, supporting a recent model indicating many genes with small effects (Zhang et al. 2015) rather than the classical model with only four genes controlling plant height (Quinby and Karper 1954). The progeny lines were much taller than either parent, suggesting that *S. bicolor* and *S. halepense* might contribute different sets of genes for plant height to their progenies, or that height in early generations may be a result of heterosis.

QTL correspondence

QTL co-localization may suggest a single gene with pleiotropic effect influencing multiple traits, or different genes that are tightly linked either by chance or due to selection during sorghum evolution. A well-known example is the *Ma1* and *Dw2* genes linked in the large heterochromatin region on chromosome 6 (Cuevas et al. 2016; Morris et al. 2013; Zhang et al. 2015). Two additional genomic locations suggest correlations between plant height and flowering time, one in the region of 6–46 Mb on chromosome 6, presumably reflecting *Ma6* and *Dw4*; and the other in the region of 56–59.5 Mb on chromosome 9 (Zhang et al. 2015). In our study, we validated these two regions on chromosome 6 and discovered four additional regions with QTL likelihood peaks at ~ 66.7 Mb on chromosome 1C, ~ 4.3 Mb and 61–64 Mb on chromosome 4A and 4D, respectively, 57–58 Mb on chromosome 7D and ~ 44.7 Mb on chromosomes 8A, conferring QTLs for **PH** and **FL** in the H6 population. High correlation between PH and FL (Fig. 1) may also contribute to the discovery of co-localized QTLs (Fig. 6).

Correspondence of QTL regions between three populations sharing *S. bicolor* BTx623 as a common parent, with the other parents being morphologically and genetically distinct genotypes that represent cultivated (IS3620C), wild diploid (*S. propinquum*) and wild polyploid (*S halepense*) sorghums, provides information about common QTLs shared between or among populations and taxon-specific QTLs that contribute to divergence (Fig. 6). Genomic regions conferring previously characterized plant height genes, *Dw1*, *Dw2* and *Dw3*, have been validated in the three sorghum populations, with both the SH-BC$_1$F$_2$ populations and the IS-RIL population segregating for all three genes, and the PQ-RIL population detecting *Dw3* (and with *Dw2* probably largely eliminated due to its close linkage to *Ma1*). Many additional QTLs for **PH** and **FL** significant in the BC$_1$F$_2$ but not in the IS-RIL or PQ-RIL populations may suggest the divergent control of plant height or inadequate statistical power to detect some QTLs in some populations. Similarly, five regions controlling **FL** corresponded in the BC$_1$F$_2$, IS-RIL and PQ-RIL populations. Six regions might harbor the known genes controlling flowering in sorghum [Table 6; (Cuevas et al. 2016; Murphy et al. 2011; Wolabu et al. 2016; Yang et al. 2014a, b)].

Despite the fact that the IS-RIL population has higher resolution and power to detect QTLs due to larger size, we detected more novel QTLs in the two SBSH- BC$_1$F$_2$ populations than any other populations, demonstrating the genetic novelty arisen during the diversification of *S. halepense*. Lack of unique QTLs in the PQ-RIL population may be mainly due to a relatively low resolution genetic map with simple sequence repeat (SSR) markers and a smaller sample size compared to the IS-RIL. Polyploidy and subsequent 'diploidization' processes are often associated with the formation of genetic novelty (Comai 2005) and the relative abundance of QTLs unique to the SH-derived populations is characteristic of this phenomenon. Noting that a genome duplication had not survived in the sorghum lineage for 96 million years before the evolution of *S. halepense*, an attractive hypothesis for further investigation is that *S. halepense* may harbor a rich collection of novel alleles that offer potential opportunities for sorghum improvement.

Acknowledgments We appreciate the support of the USDA Biotechnology Risk Assessment Program (2012-01658 to AHP and TSC), USAID Feed The Future (AID-OAA-A-13-00044 to AHP, TSC) program, and NIFA Global Food Security CAP (2015-68004-23492 to AHP, JNB). We thank members of the Plant Genome Mapping Laboratory (PGML) for help with field work.

Author contribution statement WQK conducted the experiment, analyzed the data wrote the manuscript; PN conducted the experiment and collected phenotypic data; TSC designed, supervised the experiment and edited the paper; VHG, JSR, GRP, CL and RC carried out the field experiment and collected the phenotypic data. AHP designed and supervised the experiment, edited the manuscript and recommended analytical suggestions.

Funding We appreciate the support of the USDA Biotechnology Risk Assessment Program (2012-01658 to AHP and TSC), and NIFA Global Food Security CAP (2015-68004-23492 to AHP, JNB). This work was funded in part by the United States Agency for International Development (USAID) Bureau for Resilience and Food Security under Agreement # AID-OAA-A-13-00044 (to AHP, TSC) as part of Feed the Future Innovation Lab for Climate Resilient Sorghum. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors alone.

Availability of data and material Genotypic data is available at https://www.frontiersin.org/articles/10.3389/fpls.2020.00467/full#supplementary-material. Phenotypic data can be found in the supplementary documents.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Aljanabi SM, Forget L, Dookun A (1999) An improved and rapid protocol for the isolation of polysaccharide- and polyphenol-free sugarcane DNA. Plant Mol Biol Rep 17:281

Andolfatto P, Davison D, Erezyilmaz D, Hu TT, Mast J, Sunayama-Morita T, Stern DL (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res 21:610–617

Bishop GJ (2003) Brassinosteroid mutants of crops. J Plant Growth Regul 22:325–335

Brady JA (2006) Sorghum Ma5 and Ma6 maturity genes. Texas A&M University

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890

Brown PJ, Rooney WL, Franks C, Kresovich S (2008) Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics 180:629–637

Childs KL, Miller FR, Cordonnier-Pratt MM, Pratt LH, Morgan PW, Mullet JE (1997) The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B. Plant Physiol 113:611–619

Comai L (2005) The advantages and disadvantages of being polyploid. Nat Rev Genet 6:836–846

Cuevas HE, Zhou C, Tang H, Khadke PP, Das S, Lin YR, Ge Z, Clemente T, Upadhyaya HD, Hash CT, Paterson AH (2016) The evolution of photoperiod-insensitive flowering in sorghum, a genomic model for panicoid grasses. Mol Biol Evol 33:2417–2428

Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, Shimatani Z, Yano M, Yoshimura A (2004) Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. Genes Dev 18:926–936

Fernandez MGS, Becraft PW, Yin Y, Lübberstedt T (2009) From dwarves to giants? Plant height manipulation for biomass yield. Trends Plant Sci 14:454–461

Hart GE, Schertz KF, Peng Y, Syed NH (2001) Genetic mapping of Sorghum bicolor (L.) Moench QTLs that control variation in tillering and other morphological characters. Theor Appl Genet 103:1232–1242

Higgins RH, Thurber CS, Assaranurak I, Brown PJ (2014) Multiparental mapping of plant height and flowering time QTL in partially isogenic sorghum families. G3-Genes Genom Genet 4:1593–1602

Hill CB, Li C (2016) Genetic Architecture of Flowering Phenology in Cereals and Opportunities for Crop Improvement. Front Plant Sci 7:1906

Hilley J, Truong S, Olson S, Morishige D, Mullet J (2016) Identification of Dw1, a regulator of sorghum stem internode length. PLoS ONE 11:e0151271

Hilley JL, Weers BD, Truong SK, McCormick RF, Mattison AJ, McKinley BA, Morishige DT, Mullet JE (2017) Sorghum Dw2 encodes a protein kinase regulator of stem internode length. Sci Rep 7:4616

Jung C, Müller AE (2009) Review: flowering time control and applications in plant breeding. Trends Plant Sci 14:563–573

Kong L, Dong J, Hart GE (2000) Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple-sequence repeats (SSRs). Theor Appl Genet 101:438–448

Kong W, Jin H, Franks CD, Kim C, Bandopadhyay R, Rana MK, Auckland SA, Goff VH, Rainville LK, Burow GB, Woodfin C, Burke JJ, Paterson AH (2013) Genetic analysis of recombinant inbred lines for *Sorghum bicolor* × *Sorghum propinquum*. G3 Genes Genomes Genetics 3:101–108

Kong W, Guo H, Goff V, Lee T-H, Kim C, Paterson A (2014) Genetic analysis of vegetative branching in sorghum. Theor Appl Genet 127:2387–2403

Kong W, Kim C, Zhang D, Guo H, Tan X, Jin H, Zhou C, Shuang LS, Goff V, Sezen U, Pierce G, Compton R, Lemke C, Robertson J, Rainville L, Auckland S, Paterson AH (2018) Genotyping by Sequencing of 393 Sorghum bicolor BTx623 x IS3620C recombinant inbred lines improves sensitivity and resolution of QTL Detection. G3 (Bethesda) 8:2563–2572

Kong W, Nabukalu P, Cox TS, Goff VH, Pierce GJ, Lemke C, Robertson JS, Compton R, Tang H, Paterson AH (2020) Transmission genetics of a *Sorghum bicolor* × *S. halepense* backcross populations. Front. Plant Sci 11:467

Ku LX, Zhang LK, Tian ZQ, Guo SL, Su HH, Ren ZZ, Wang ZY, Li GH, Wang XB, Zhu YG, Zhou JL, Chen YH (2015) Dissection of the genetic architecture underlying the plant density response by mapping plant height-related traits in maize (*Zea mays* L.). Mol Genet Genomics 290:1223–1233

Kutschera U, Wang ZY (2012) Brassinosteroid action in flowering plants: a Darwinian perspective. J Exp Bot 63:3511–3522

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using Rflp linkage maps. Genetics 121:185–199

Li X, Li XR, Fridman E, Tesso TT, Yu JM (2015) Dissecting repulsion linkage in the dwarfing gene Dw3 region for sorghum plant height provides insights into heterosis. Proc Natl Acad Sci USA 112:11823–11828

Lin YR, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics 141:391–411

McWhorter C (1971) Introduction and spread of johnsongrass in the United States. Weed Sci, pp 496–500

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci USA 110:453–458

Multani DS, Briggs SP, Chamberlin MA, Blakeslee JJ, Murphy AS, Johal GS (2003) Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. Science 302:81–84

Murphy RL, Klein RR, Morishige DT, Brady JA, Rooney WL, Miller FR, Dugas DV, Klein PE, Mullet JE (2011) Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. Proc Natl Acad Sci USA 108:16469–16474

Murphy RL, Morishige DT, Brady JA, Rooney WL, Yang SS, Klein PE, Mullet JE (2014) Ghd7 (Ma(6)) Represses sorghum flowering in long days: Ghd7 alleles enhance biomass accumulation and grain production. Plant Genom 7:1–10

Murray SC, Rooney WL, Mitchell SE, Sharma A, Klein PE, Mullet JE, Kresovich S (2008) Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates. Crop Sci 48:2180–2193

Murray SC, Rooney WL, Hamblin MT, Mitchell SE, Kresovich S (2009) Sweet sorghum genetic diversity and association mapping for brix and height. Plant Genome-Us 2:48–62

Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol 4:133–142

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Lyons E, Maher C, Narechania A, Penning B, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein PE, Kresovich S, McCann MC, Ming R, Peterson DG, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang ZW, Millard MJ, Gardner CAC, McMullen MD, Holland JB, Bradbury PJ, Buckler ES (2014) The genetic architecture of maize height. Genetics 196:1337

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) Green revolution' genes encode mutant gibberellin response modulators. Nature 400:256–261

Quinby J (1966) Fourth maturity gene locus in sorghum. Crop Sci 6:516–518

Quinby J, Karper R (1945) Inheritance of three genes that influence time of floral initiation and maturity date in milo. J Am Soc Agron 37(11):916–936

Quinby JR, Karper RE (1954) Inheritance of Height in Sorghum. Agron J 46:211–216

Quinn LD, Barney JN, McCubbins JS, Endres AB (2013) Navigating the "noxious" and "invasive" regulatory landscape: suggestions for improved regulation. Bioscience 63:124–131

R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna

Ritter KB, Jordan DR, Chapman SC, Godwin ID, Mace ES, McIntyre CL (2008) Identification of QTL for sugar-related traits in a sweet x grain sorghum (*Sorghum bicolor* L. Moench) recombinant inbred population. Mol Breed 22:367–384

Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M (2002) Green revolution: a mutant gibberellin-synthesis gene in rice. Nature 416:701–702

Sezen UU, Barney JN, Atwater DZ, Pederson GA, Pederson JF, Chandler JM, Cox TS, Cox S, Dotray P, Kopec D, Smith SE, Schroeder J, Wright SD, Jiao Y, Kong W, Goff V, Auckland S, Rainville LK, Pierce GJ, Lemke C, Compton R, Phillips C, Kerr A, Mettler M, Paterson AH (2016) Multi-phase US spread and habitat switching of a post-columbian invasive *Sorghum halepense*. Plos One 11:e0164584

Upadhyaya HD, Wang Y-H, Sharma S, Singh S (2012) Association mapping of height and maturity across five environments using the sorghum mini core collection. Genome 55:471–479

Wang YH, Li JY (2006) Genes controlling plant architecture. Curr Opin Biotech 17:123–129

Wang X, Wang J, Jin D, Guo H, Lee TH, Liu T, Paterson AH (2015) Genome alignment spanning major poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. Mol Plant 8:885–898

Wang Y, Zhao J, Lu W, Deng D (2017) Gibberellin in plant height control: old player, new story. Plant Cell Rep 36:391–398

Warnes GR, Bolker B, Bonebakker L, Gentleman R, LIaw W, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B (2016) gplots: Various R programming tools for plotting data 2(4):1

Wolabu TW, Zhang F, Niu LF, Kalve S, Bhatnagar-Mathur P, Muszynski MG, Tadege M (2016) Three FLOWERING LOCUS T-like genes function as potential florigens and mediate photoperiod response in sorghum. New Phytol 210:946–959

Yamaguchi M, Fujimoto H, Hirano K, Araki-Nakamura S, Ohmae-Shinohara K, Fujii A, Tsunashima M, Song XJ, Ito Y, Nagae R, Wu J, Mizuno H, Yonemaru J, Matsumoto T, Kitano H, Matsuoka M, Kasuga S, Sazuka T (2016) Sorghum Dw1, an agronomically important gene for lodging resistance, encodes a novel protein involved in cell proliferation. Sci Rep 6:28366

Yang SS, Murphy RL, Morishige DT, Klein PE, Rooney WL, Mullet JE (2014a) Sorghum phytochrome b inhibits flowering in long days by activating expression of SbPRR37 and SbGHD7, Repressors of SbEHD1, SbCN8 and SbCN12. Plos One 9(8):e105352

Yang SS, Weers BD, Morishige DT, Mullet JE (2014b) CONSTANS is a photoperiod regulated activator of flowering in sorghum. BMC Plant Biol 14(1):1–5

Zhang D, Guo H, Kim C, Lee TH, Li JP, Robertson J, Wang XY, Wang ZN, Paterson AH (2013) CSGRqtl, a comparative quantitative trait locus database for saccharinae grasses. Plant Physiol 161:594–599

Zhang D, Kong W, Robertson J, Goff VH, Epps E, Kerr A, Mills G, Cromwell J, Lugin Y, Phillips C, Paterson AH (2015) Genetic analysis of inflorescence and plant height components in sorghum (Panicoidae) and comparative genetics with rice (Oryzoidae). BMC Plant Biol 15:107